

*Матеріали IV Всеукраїнської науково-технічної конференції ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РАДІОТЕХНІКИ, ПРИЛАДОБУДУВАННЯ І КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ 2019*

УДК 510.589

**Михайло Паламар д.т.н., проф., Михайло Стрембіцький к.т.н., Андрій Чайковський к.т.н., доц., Юрій Пастернак, Володимир Круглов**

Тернопільський національний технічний університет імені Івана Пулюя, Україна

**НАВЧАННЯ ЗГОРТАЛЬНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПОБУДОВИ СИСТЕМИ КОМП'ЮТЕРНОГО ЗОРУ**

Запропоновано спосіб використання згортальних нейронних мереж для побудови системи розпізнавання зображення. Розглянуто класичні способи розпізнавання зображення. Наведено переваги використання згортальних нейронних мереж для побудови системи комп'ютерного зору.

Ключові слова: нейронна мережа, алгоритм навчання, комп'ютерний зір.

**Mykhaylo Palamar, Mykhaylo Strembitskiy, Andrii Chaikovskiy, Yuriy Pasternak, Volodymyr Kruglov**

**STUDYING CLUSTER NEURAL NETWORKS FOR COMPUTER VISION**

A method of using crimping neural networks for constructing an image recognition system is proposed. Classical ways of image recognition are considered. The advantages of using the curvilinear neural network for constructing a computer vision system are presented.

Keywords: neural network, learning algorithm, computer vision.

**Вступ.** Розпізнавання зображень, рукописних символів, картинок доволі важко формалізувати [1, 2]. Щодо розпізнавання рукописних символів, то існує база даних MNIST, яка містить 60000 навчаючих впар (зображення-мітка) і 10000 тестових (зображення без мітки). Усі зображення нормалізовані по розміру і відцентровані. Зокрема саме для розпізнавання тексті і зображень використовується система комп'ютерного зору.

**Мета роботи:** створення і навчання нейронної мережі для розпізнавання рукописних символів, приймаючи зображення на вході та активуючи один із 10 виходів.

У нейронній мережі прямого поширення сигналу кожний нейрон зв'язаний з нейронами наступного і попереднього шару. Сигнал проходить лише в напрямку від вхідного шару до вихідного без використання рекурсії або ліній затримки [3]. Спочатку необхідно вирішити завдання яким чином подавати дані на вхід. Найпростіший спосіб і майже безальтернативний – це виділити двохмірну матрицю зображення і подати її у вигляді одномірного вектора. Тобто для зображення розміром 28x28 потрібно використати 784 входів, що є не дуже мало.

Існує багато різних методик для вибору параметрів нейронної мережі, одна із таких методик стверджує, що кількість нейронів проміжного шару повинна бути хоча б на порядок більшою за кількість входів. Приймаючи такі твердження нейронна мережа матиме порядку 15000 нейронів.

**Згортальні нейронні мережі.** Вирішення проблеми великої кількості нейронів було знайдено американським ученим французького походження Яном ЛеКуном, він запропонував використовувати так звані загортальні нейронні мережі [4, 5].

Ідея загортальних нейронних мереж полягає в чергуванні загортальних шарів (C-layers), субдискретизуючих шарів (S-layers) і наявності повнозв'язних (F-layers) шарів на виході [6, 7].



Рис. 1. Архітектура загортальної нейронної мережі

Така архітектура включає в себе 3 основних парадигми:

1. Локальне сприйняття.
2. Розділювальні ваги.
3. Субдискретизація.

Локальне сприйняття полягає в тому, що на вхід одного нейрона подається не все зображення (або вихід попереднього шару), а лише деяка ділянка. Такий підхід дозволяє зберегти топологію зображення від шару до шару.

Концепція розділених ваг передбачає, що для більшої кількості зв'язків використовується не великий набір вагових коефіцієнтів. Тобто для вхідного зображення розмірами 32x32 пікселі кожен з нейронів наступного шару прийме лише невелику ділянку розміром наприклад 5x5, причому кожен з фрагментів буде опрацьований одним і тим же набором [8]. Самих наборів ваг може бути багато, однак кожен з них буде застосовуватися до всього зображення. Такі набори називаються ядрами (kernels). Тоді для 10 ядер розміром 5x5 для вхідного зображення розмірами 32x32 кількість зв'язків буде дорівнювати 256000, а кількість налаштовуваних параметрів лише 250.

Такий підхід для розпізнавання дає кращі результати, оскільки штучно введено обмеження на ваги покращує узагальнюючі властивості мережі (generalization), що у підсумку позитивно відображається на властивості мережі знаходити інваріанти в зображенні і реагувати головним чином на них, не приймаючи до уваги інший шум. При класичному розпізнаванні зображень системи будуються на основі двохмірних фільтрів. Фільтр являє собою матрицю коефіцієнтів, зазвичай задану вручну. Така матриця застосовується до зображення з допомогою математичної операції згортки.

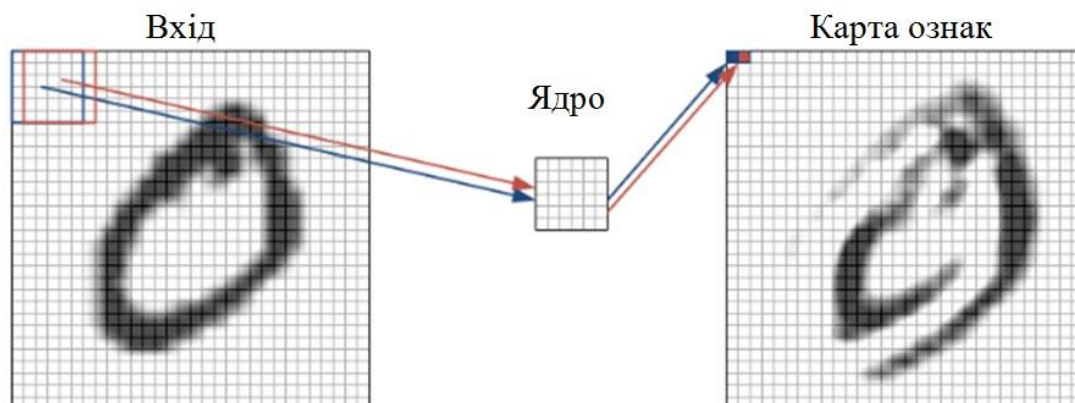


Рис. 2. Фрагментування зображення при використанні ядра

Суть такої операції в тому, що кожний фрагмент зображення перемножується на матрицю (ядро) згортки поетапно і результат додається і записується в аналогічну позицію вихідного зображення. Образ зображення згорнутий з деяким ядром дасть інше зображення, кожний піксель якого буде означати ступінь схожості фрагмента зображення на фільтр. Тобто отримаємо карту ознак.

Кожен фрагмент зображення поелементно перемножується на невелику матрицю ваг (ядро), результат додається. Ця сума є пікселем вихідного зображення, яке називається картою ознак. Зважена сума входів ще пропускається через функцію активації (що є обов'язковим для будь-якої нейронної мережі), це може відбуватися і в S-шарі. Варто відмітити, що в ідеалі не різні фрагменти проходять послідовно через ядро, а паралельно все зображення проходить через ідентичні ядра. Окрім цього кількість ядер (наборів ваг) визначається розробником і залежить від того яку кількість ознак необхідно виділити. Ще одна особливість загорткового шару в тому, що він дещо зменшує зображення за рахунок крайових ефектів. Суть субдискретизації та S-шарі полягає в зменшенні просторової розмірності зображення. Тобто вхідне зображення зменшується на задану кількість раз. Частіше за все в 2 рази, хоча може бути і не рівномірне зменшення, наприклад, 2 по вертикалі та 3 по горизонталі. Субдискретизація потрібна для забезпечення інваріантності до масштабу.

Чергування шарів дозволяє складати карти ознак і карти подібностей, що на практиці означає властивість розпізнавати складні ієрархічні ознаки. Після проходження декількох шарів карта ознак вироджується в вектор або навіть скаляр, однак таких пар ознак формується сотні. В такому вигляді вони подаються на один-два шари повнозв'язної нейронної мережі. Вихідний шар такої мережі може мати різні функції активації. В простому випадку може бути тангенціальна функція, також успішно використовуються радіальні базисні функції.

**Навчання загортальної нейронної мережі.** Для навчання нейронної мережі потрібно визначити як буде змінюватися якість розпізнавання. В нашому випадку для цього було використано поширену в теорії нейронних мереж функцію середньоквадратичної похибки [2].

$$E^p = \frac{1}{2} (D^p - O(I^p, W))^2 \quad (1)$$

де  $E_p$  – похибка розпізнавання для  $p$ -тої навчаючої пари;

$D_p$  – бажаний вихід мережі;

$O(I_p, W)$  – вихід мережі, який залежить від  $p$ -того входу і вагових коефіцієнтів  $W$ , куди входять ядра згортки, зміщення, вагових коефіцієнтів  $S$  - та  $F$  - шарів.

Завдання навчання полягає у налаштуванні ваги  $W$  таким чином, щоб вони для будь-якої навчальної пари  $(I_p, D_p)$  давали найменшу похибку  $E_p$ . Розрахунок похибки для всіх навчальних вибірок, відбувається вибором середньо арифметичного по похибках для всіх навчальних пар. Таку усереднену похибку позначимо як  $E$ .

Для мінімізації функції помилки  $E_p$  самими ефективними є градієнтні методи. Якщо розкласти в ряд Тейлора функцію помилки  $E_p$ , то отримаємо наступний вираз:

$$E(W) = E(W_c) + (W - W_c) \frac{dE(W_c)}{dW} + \frac{1}{2} (W - W_c)^2 \frac{d^2E(W_c)}{dW^2} + \dots \quad (2)$$

де  $E$  – функція помилок;

$W_c$  – деяке початкове значення ваг.

Для знаходження максимуму функції візьмемо похідну помилки по вагах, не враховуючи члени вище 2-го порядку:

$$\frac{dE(W)}{dW} = \frac{dE(W_c)}{dW} + (W - W_c) \frac{d^2E(W_c)}{dW^2} \quad (3)$$

З цього вираз отримаємо, що ваги, при яких значення функції помилки буде мінімальним вирахуємо з наступного виразу:

$$W_{min} = W_c - \left( \frac{d^2 E(W_c)}{dW^2} \right)^{-1} \frac{dE(W_c)}{dW} \quad (4)$$

Тобто оптимальне вагове значення обчислюється як поточне значення мінус похідна функції помилки по ваговому значенню, розділена на другу похідну від функції помилки. Для багатовимірного випадку (тобто для матриці ваг) все так само, лише перша похідна перетворюється в градієнт (вектор часткових похідних), а друга похідна перетворюється в Гесіан (матрицю других часткових похідних).

І тут виділимо два варіанти. Якщо опустити другу похідну, то отримаємо алгоритм найшвидшого градієнтного спуску. Гесіан замінимо простішим Левенберга-Марквардта (ЛМ), його апроксимацією квадратним якобіаном.

Алгоритм ЛМ вимагає обробки всієї навчальної вибірки, тоді як алгоритм градієнтного спуску може працювати з кожною окремо взятою навчальною вибіркою. В останньому випадку алгоритм назвемо стохастичним градієнтом. З огляду на, що база містить 60000 навчальних зразків нам більше придатним буде стохастичний градієнт. Ще однією перевагою стохастичного градієнта є його менша схильність потрапляння в локальний мінімум в порівнянні з алгоритмом ЛМ.

**Висновок.** Представлені формули дозволяють легко обчислити похідну помилки для вагових коефіцієнтів нейронної мережі. Розглянутий підхід для розпізнавання зображень дає кращі результати, порівняно із традиційними методами, оскільки штучно введено обмеження на вагові коефіцієнти покращує узагальнюючі властивості мережі, що позитивно відображається на властивості мережі знаходити подібні зображення, навіть в умовах дії шумів.

### Література

1. Charles C. Tappert, Sung-Hyuk Cha: English Language Handwriting Recognition Interfaces. Text Entry Systems, ed. MacKenzie and Tanaka-Ishii, Morgan Kaufman, 2007.
2. Степашко П.В. Огляд наявних підходів до розв'язання задачі розпізнавання рукописного тексту // Індуктивне моделювання складних систем, № 5, 2013, м. Київ – с. 278 – 287.
3. Zhang G.P.: Neural networks for classification: a survey. IEEE Transactions on Systems, Man, and Cybernetics, 2000 - Part C: Applications and Reviews, 30(4):451-462.
4. Y. LeCun and Y. Bengio: Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, MIT Press, 1995.
5. Y. LeCun, L. Bottou, G. Orr and K. Muller: Efficient BackProp, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998.
6. Nielsen M. [Електронний ресурс]: Neural Networks and Deep Learning. – 2017. – Режим доступу: <http://neuralnetworksanddeeplearning.com/chap2.html>.
7. Семенов С. Г. Інтелектуальна система контролю стану небезпечних ділянок залізничного шляху / С. Г. Семенов, О. В. Ліпчанська // Національний технічний університет «Харківський політехнічний інститут». Сучасні інформаційні системи. Т. 2, № 2, 2018., Харків – с. 89 – 93.
8. Стрембіцький М. О. Selection of the efficient video data processing strategy based on the analysis of statistical digital images characteristics / Михайло Паламар, Мирослава Яворська, Михайло Стрембіцький, Володимир Стрембіцький // Вісник Тернопільського національного технічного університету, - Тернопіль, 2018. № 3 (91) – С. 107-114.